Marquette University
e-Publications@Marquette

Master's Essays (1922-2009) Access Restricted to Marquette Campus

Dissertations, Theses, and Professional Projects

4-1-1979

# The Effects of Criterion Based Evaluation on Reliability and Self-Evaluation in a Preclinical Dental Laboratory

William Keith Duncan
*Marquette University*

# THE EFFECTS OF CRITERION BASED EVALUATION ON RELIABILITY AND SELF-EVALUATION IN A PRECLINICAL DENTAL LABORATORY

by

William Keith Duncan, D.D.S.

An Essay submitted to the Faculty of the Graduate School, Marquette University, in Partial Fulfillment of the Requirements for the Degree of Master of Education

Milwaukee, Wisconsin
April, 1979

TABLE OF CONTENTS

LIST OF TABLES

Dentistry has always been considered an art as well as a science by those who practice it and by those who teach it. The scientific aspects include gathering data, making diagnoses, and formulating treatment plans. Each of these aspects is relatively systematic and objective to the practitioner. However the implementation, the act of preparing and restoring a cavity in a tooth, is largely art. Evaluation of art is a very subjective process involving individual judgments based on the predispositions of the evaluator.

Traditionally students have received a letter grade for the psychomotor task of preparing and restoring a cavity in a tooth. If any feedback was given it generally reflected the instructor's personal preferences. (Yates, 1976). Instructors frequently have been taught, or have acquired, widely different evaluation procedures and criteria. The glance and grade system in which an overall grade is assigned encourages subjectivity and fails to inform the student of specific strengths and weaknesses. Many studies have shown that agreement among raters of cavity preparations and restorations is low. (Natkin, 1967) (Gaines, 1974). Surveys have also shown that inconsistent faculty evaluation is a significant source of discouragement as well as the major reason for the student decision to do just enough to get by. (Natkin, 1967). The consequences of behavior are probably the most important determinants of what is learned and of the efficiency of learning. When the consequences of behavior are reliable and consistent, learning tends to be predictable and efficient. Conversely, when they are inconsistent, learning tends to be unpredictable and inefficient. Therefore, efforts have been made to improve the reliability of faculty evaluations of student performance of psychomotor tasks.

Specific types of cavity designs were described by G.V. Black as early as 1920. Over the years these basic forms have been used, with some modification, to judge the general quality of student cavity preparations. However, dental educators are in agreement that procedures in dentistry are complex in nature, depending upon the proper completion of the component parts in order to achieve an acceptable result. If teaching is to be effective, students should be evaluated on each of the component parts of the task utilizing specific criteria for each discrete step. Benefits should be derived from stating the specific criteria to be met by the students and by requiring raters to judge preparations on each criterion. Several studies (Natkin, 1967) (Gaines, 1974) (Ryge, 1973) (Houpt, 1973) have shown that precise definition has brought about higher agreement among raters.

Furthermore, in order to become qualified practitioners, dental students must learn the criteria for ideal products and be abīe to judge when these criteria have been met. It appears there is an assumption that the practicing dentist possesses this crucially important ability to appraise his/her own skills without receiving formal training, since this attribute or skill is rarely considered in the dental curriculum.

## Literature Review

After reviewing the literature Irion, (1965) concluded that feedback is the most important variable governing skill learning. Nedelsky, (1965) demonstrated that unfair or inconsistant systems of evaluation will negatively affect the morale of students.

Natkin, (1967) demonstrated that the evaluations of student performances are highly inconsistent. He also stated that few clinical departments have formulated explicit written criteria for evaluation of student performance. He showed that instructors using a ten-point scale ranging from A-E to evaluate students' preclinical endodontic procedures arrived at a grade of 3.3 as a mean range of grades; and in 45% of the cases the range exceeded 4.0 grade points. Fuller (1972) found no significant agreement between instructors using the traditional "glance and grade" system of evaluation, which is commonly used in many dental school departments and many state board exams. Houpt (1973 reported that instructors evaluating cavity preparations for a second time change only 20% of the ratings. The results of the study showed that evaluation of total clinical projects are relatively reliable and accurate; but when raters are asked to evaluate individual criteria the reliability is low (.20). Hinkelman (1973) states that once it has been decided what criteria are to be used in evaluation, the level at which the student achieves the criteria becomes the next problem. He recommends a two-point, or pass-fail system in which the student either achieved the criteria or failed to achieve the criteria as demonstrated in Houpt's study. Howevér, Hinkelman goes on to say, "A two-point system is not practical in preclinical learning. The problem arrises as to which of the criteria should be regarded as passing, the ideal or the clinically acceptable." O'Conner recommends the use of comparison stimuli as aids for rating projects. In her study photographs of preparations representing "adequate" attainment (grade of 3 on a 1-5 scale) of each of six criteria were used as

references in rating all preparations. Her results were not definitive but she concludes that the use of exemplars shows promise for improving inter-rater agreement.

Abrams (1974), in his article on self evaluation states that there can often be a problem of interpretation of criteria between the student and instructor. It is implicit that both agree as to definitions and the meaning of criteria. Often the student completes a procedure incorrectly as a result of misinterpreted criteria.

The purpose of this study was two-fold. One was to establish a high consistency of evaluation among raters of dental psychomotor skills. Higher consistency in feedback should lead to improved learning and less frustration of the student. This study included establishing an operationally defined criteria check-list for specific pre-clinical psychomotor tasks. It was hypothesized that utilization of such a check-list would result in relatively high inter-rater reliability estimates. Students and faculty were familiarized with the specific criteria and all evaluation was based on them.

The second purpose was to increase students' ability to self-evaluate their projects. The practicing dentist must be able to recognize and apply criteria of acceptability to his/her products. Students were asked to evaluate their own products on the same criteria check-list as the faculty. It was hypothesized that students who knew the specific criteria would perform better on the average than those who did not. If student's evaluation of his/her own projects closely corresponded to faculty evaluations, he/she received a bonus point toward the course grade. Specifically, it was hypothesized that students who received a bonus point would have higher cumulative

project points (excluding bonus points) than those students who did not receive a bonus point.

## Methodology

### Subjects:

The Pedodontic Department of the Marquette University School of Dentistry has as a part of its curriculum a sophomore preclinical laboratory. The subjects who participated in this study were dental students enrolled in the course. There was no selection process, everyone in the class of 138 participated. These students range in age from 21 to 26 years. There were 12 females and 126 males. The subjects were broken into two groups of 69 by alphabetical order to facilitate the evaluation process. The subjects had had previous experience preparing teeth in other courses but had no experience with criteria grading or self evaluation while in dental school.

### Apparatus:

The apparatus used by the students in preparing and restoring the teeth was standard dental equipment. Teeth were prepared using Starr Futura highspeed and Midwest lowspeed handpieces containing a #330 carbide burr. Restorations were completed using a Dispersalloy prepackaged amalgam in posterior cavities and Adaptic resin in anterior cavities. The techniques and instrumentation used in placing, carving, and polishing the restorations were standard dental procedures. The cavities were prepared on a Columbia Pedodontic Typodont which utilizes ivorine plastic teeth.

Three criteria checklists were specifically developed for this study. The procedure followed was that the anterior and posterior cavities were divided into two major areas; the preparation phase and the restoration phase. These phases were then divided into their essential component parts. An operationally defined criteria describing the psychomotor task to be completed in each component was then written. The criteria checklist for posterior teeth contained eight criteria for the preparation phase and eight criteria for the restoration phase (Appendix A). The checklist for the anterior teeth contained five criteria for the preparation phase and six criteria for the restoration phase (Appendix B). The stainless steel crown restoration was also divided into a preparation and restoration phase and criteria were developed. The stainless steel crown checklist contains four criteria in the preparation phase and five criteria in the restoration phase (Appendix C). These criteria were developed by the full time faculty of the department of Pedodontics in consultation with the school's educational psychologist. The rating scale for each criteria was simply a pass-fail system in which the student either achieved or did not achieve the specific criteria. Philosophically the department of Pedodontics feels that the "ideal" is the goal in a preclinical laboratory. The term "clinically acceptable" has no place in the laboratory. Realistically, an individual student will not attain the "ideal" for each of the criteria but certainly this should be the goal and the yardstick for evaluation, particularly under the controlled environment of the laboratory.

Procedure:

In lectures prior to the inception of the laboratory, students were given instructions on its operation. The criteria to be met for each project were described and discussed. These criteria were also listed in the students' laboratory manual. The instructors in the preclinical laboratory were given one orientation session to expose them to the criteria which had been established and how they would be used.

At the first laboratory session the students were told that their projects would be evaluated objectively, based on the criteria which had been established. The laboratory consisted of practice sessions and practical examinations. During the practice sessions all projects were critiqued using the established criteria. Following the practice sessions a practical examination was given. For the practical the student was asked to prepare and restore cavities on two posterior teeth and two anterior teeth. The student was instructed to determine which of the two posterior preparations were most ideal and to maintain it for evaluation. He/she was instructed to restore the other preparation for evaluation of the restoration. The instructions were identical for anterior teeth; that is, maintain the most ideal preparation and restore the other. The students were asked to evaluate their own preparation and restoration utilizing the criteria system. The students were given two hours to complete the practical examination and were then asked to leave the laboratory.

Two teams of three instructors each were established to evaluate the projects. Team A evaluated students 1-69 on practical (1) and team B evaluated students 69-138. On the second practical the two teams

evaluated the opposite halves of the class. Each team consisted of one full time faculty member and two graduate students in Pedodontics. Each member of the team evaluated the projects independently. Therefore, for each of the four projects on practical 1 (anterior preparation, anterior restoration, posterior preparation, posterior restoration) the students received three grades from instructors plus the self-evaluation grade. The numerical values of these grades were determined as follows. The preparation of the tooth was evaluated on eight criteria. A number grade was given from 1-4 depending upon the number of criteria which were met.

| Criteria Met | Number Grade |
|---|---|
| 8 | 4 |
| 7 | 3 |
| 6 | 2 |
| 5 | 1 |
| less than 5 | 0 |

The restoration of the tooth also being evaluated on eight criteria was assigned a number grade in the same manner. Each instructor therefore assigned a number grade of 0-4 for each of the student's projects.

For practical (2), two teeth were prepared to receive a stainless steel crown. The most ideal preparation, as determined by the student was maintained for evaluation and the other was fitted with the crown for evaluation of the finished restoration. As before, the instructor assigned a number grade, from 0-4, based on the criteria which were met.

The data from the evaluation was divided into groups based on the task and the evaluation team. A total of 12 subsets were analyzed. These included posterior preparation, posterior restoration, anterior

preparation, anterior restoration, crown preparation, and crown restoration for each of the two evaluation teams. A treatment by subjects design was chosen to analyze the data. The data was run on the Marquette Sigma 9 computer utilizing the Vanderbilt Statistical Package Program for the chosen design. The Scheffe post hoc test was used to determine the presence of significant differences between specific pairs of evaluators. Finally, the reliability of the evaluators was investigated using the method for single-factor experiments with repeated measures as described by Winer (1971).

Self-evaluation scores were multiplied by three and this total was compared to the total of the three instructor evaluations. If these two totals were within 1 point of each other the student was awarded a bonus point toward his/her total. The scores (exclusive of bonus point) of students receiving the bonus point were compared with those who did not by a t-test. The hypothesis being that these students who knew and understood the criteria would be better able to critique their work and also better prepared to perform the task.

## RESULTS

The results demonstrating the consistency of evaluation are given in tables 1A through 4B. For each of the tables, evaluator number 1 was a full time faculty member and evaluator numbers 2 and 3 were pedodontic graduate students.

The analysis of variance summaries demonstrated that at the $P < .05$ level eight out of twelve samples were significantly different. At the $P < .01$ level seven out of twelve were still significant. Those samples (tables) not showing significant differences are 1D, 2A, 2C, 3A and 3B. Three of the five samples not showing significant differences

were from evaluation group A and two were from evaluation group B. Three of these samples evaluated preparations and two evaluated restorations.

The Scheffé test is an a posteriori method that is quite conservative in nature. It was used to test differences between pairs of evaluators. The five tables showing no significant differences in the analysis of variance obviously demonstrated no significance between pairs of evaluators. In the remaining seven tables twelve of the twenty-one pairs of evaluations were significantly different at the $P < .05$ level and all but one of these were significant at the $P < .01$ level. Of the twelve pairs of evaluations found to be significantly different, ten of them involved the full time faculty member and one graduate student as the evaluators. Only two were between two graduate students. The mean evaluations by faculty were consistently lower than graduate students. The results demonstrated that in eight of twelve samples the faculty person evaluated lower than the graduate students.

In the summary of reliability section of the tables, $r_1$ is the estimate of the reliability of a single evaluator. $r_3$ is the estimate of the reliability of the mean of the three evaluators. The range of $r_1$ over the twelve samples was from .29 to .55 with a mean of .43. The range of $r_3$ was from .55 to .78 with a mean of .68. When comparing the means of $r_1$ and $r_3$ for the two evaluation groups, they were found to coincide with the overall means of .43 and .68.

The results of the self-evaluation portion of this study are given in table 5. A T test compared the scores of those students who accurately evaluated their projects (as previously described) to those who did not. On the first practical the students receiving the bonus point (more

accurate evaluations) had a significantly higher mean at the $P < .01$ level. In the second practical, those students who did not receive the bonus point had a significantly higher mean at the $P < .05$ level.

## Discussion and Conclusions

The first objective of this study was to establish a high consistency of evaluation among raters of dental psychomotor tasks. The establishment of a criteria based system should be an improvement when compared to the overall grade method. One advantage is that instructors are required to focus on specific aspects of each project which are consistent regardless of the individual instructor. The establishment of these criteria provide a framework from which the instructor can build appropriate feedback for the student. This in itself should add to the reliability and consistency of the feedback. However, the results demonstrate that the inter-rater reliability is still less than hoped for. The analysis of varience points to a number of significant differences in evaluations. The scheffé test shows that in almost every case the significant difference was between a full time faculty member and a graduate student. Because of financial and time restrictions it is impossible to utilize only full time faculty in the preclinical laboratory. Therefore graduate students studying the specialty of children's dentistry are used as instructors. The graduate students have very little experience in evaluating and scoring psychomotor projects. It is recommended that an effort be made to fill this void prior to the laboratory experience. A method of doing this would be through training sessions. Training sessions should not only include discussion of the criteria to be utilized but should provide the

opportunity to apply them. Several preparations should be fabricated demonstrating both success and common errors in meeting the criteria as described. Each of the graduate students should evaluate and score these projects to improve their reliability and consistency.

A second recommendation which should improve the reliability and consistency of the feedback to students would be to change the methodology of scoring the projects on the practicals. The method of scoring could be improved by giving each criteria the weight of one point. Therefore, if the student in preparing a posterior cavity preparation successfully met six of the eight criteria he/she would receive a score of six rather than two. Some form of the criteria list should also be used to record the scores at the time of the evaluation rather than relying on memory and simply recording the score. This procedure would enable the student to know which of the criteria were not met on the practical examination. In the laboratory the students received the most meaningful feedback from the practice sessions where the criteria were discussed and explanations were given as to why certain criteria were met and others were not. Suggestions for improvement were made and the student was often asked to repeat the task. The practical itself only provided the student with a number from one to four from each instructor for each portion of the exam. The previous recommendations would make the evaluation of the practical consistent with the practice sessions of the laboratory and also would provide more meaningful feedback from the practical itself. In addition to the above described advantages, giving each criteria the weight of one point would also increase the range of scores received by the students. This effect would have two advantages, one being

that the broader range would provide a more accurate determination of each students relative performance. Secondly, this broader range would make the process of self-evaluation more meaningful, which will be discussed further later in this paper.

The second objective of this project was to expose the students to the process of self-evaluation. The hypothesis being that in order to accurately evaluate their own work the students would have to know the established criteria and that knowing the criteria would improve their performance. The students were motivated to know and understand the criteria in two ways. First, the student was given the opportunity to select from two projects the one that he/she judged most ideal for scoring purposes. Therefore it was personally beneficial for the student to be able to select the better project. The second motivating factor was the bonus point which was awarded to those students who could evaluate their projects within a certain range of the faculty evaluation. On the first practical the hypothesis seemed to be born out. There was a significantly better performance demonstrated by those students who evaluated their projects most accurately and presumably had a better understanding of the criteria. However, on the second practical, this trend was not continued, on the contrary, it was reversed. Those students who received the bonus point scored significantly lower than those who did not. It was felt that initially the students were making a sincere effort to evaluate their work but later realized that by simply selecting a middle range grade of (3) the "odds" would be in favor of receiving the bonus point. It was felt that the expanded range which would be created by scoring each criteria would dissuade students from "playing the odds" and to encourage them to make an

honest appraisal of their project.

The following conclusions were drawn from this study.

1) If graduate students are to be utilized as instructors in preclinical laboratories, they must receive an in depth orientation including practice sessions.
2) Further research is necessary in establishing high inter-rater reliabilities and consistent feedback for dental psychomotor tasks.
3) Motivating students to self-evaluate their projects can be accomplished through asking them to select one of two projects for scoring.
4) When asking students to self-evaluate projects a wide enough range must be provided to minimize meaningless guessing.

TABLE 1A

Summary of Results for Posterior
Preparations from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 194.00 | 213.00 | 218.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.812 | 3.087 | 3.159 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Columns (A) | 2 | 4.6472 | 2.3236 | 7.2881 | .0013 |
| Rows (Subs) | 68 | 59.9192 | .8812 | 2.7640 | .000 |
| Interaction | 136 | 43.3567 | .3188 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 4.33 $P < .05$
Evaluators 1-3 . . . . F = 6.83 $P < .01$
Evaluators 2-3 . . . . F = .28 not sig.

Summary of Reliability

$\theta = .58$
$r_1 = .37$
$r_3 = .64$

TABLE 1B

Summary of Results for Posterior Fillings from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 197.00 | 218.00 | 245.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.855 | 3.159 | 3.551 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Columns | 2 | 16.7825 | 8.3912 | 40.9212 | .000 |
| Rows (Subs) | 68 | 42.9824 | .6321 | 3.0825 | .000 |
| Interaction | 136 | 27.8879 | .2051 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 7.5 P<.01
Evaluators 1-3 . . . . F = 40.0 P<.01
Evaluators 2-3 . . . . F = 12.6 P<.01

Summary of Reliability

$\theta$ = .69
$r_1$ = .41
$r_3$ = .67

TABLE 1C

Summary of Results for Anterior
Preparations from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 188.00 | 211.00 | 202.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.725 | 3.058 | 2.928 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Columns | 2 | 3.8933 | 1.9467 | 7.0706 | .0016 |
| Rows (Subs) | 68 | 68.7314 | 1.0108 | 3.6712 | .000 |
| Interaction | 136 | 37.4431 | .2753 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 7.25 P<.01
Evaluators 1-3 . . . . F = 3.13 not sig.
Evaluators 2-3 . . . . F = 1.06 not sig.

Summary of Reliability

$\theta$ = .89
$r_1$ = .47
$r_3$ = .72

TABLE 1D

Summary of Results for Anterior
Fillings from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 256.00 | 260.00 | 254.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 3.710 | 3.768 | 3.681 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Columns | 2 | .2698 | .1349 | .8190 | .4466 |
| Rows (Subs) | 68 | 25.0801 | .3688 | 2.2394 | .0001 |
| Interaction | 136 | 22.3992 | .1647 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = .36 not sig.
Evaluators 2-3 . . . . F = .08 not sig.
Evaluators 2-3 . . . . F = .79 not sig.

Summary of Reliability

$\theta$ = .41
$r_1$ = .29
$r_3$ = .55

TABLE 2A

Summary of Results for Posterior
Preparations from Evaluation Group B

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 184.00 | 180.00 | 191.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.667 | 2.609 | 2.768 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | .8982 | .4491 | 1.3954 | .2500 |
| Rows (Subs) | 68 | 68.2871 | 1.0042 | 3.1202 | .000 |
| Interaction | 136 | 43.7715 | .3218 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = .18 not sig.
Evaluators 1-3 . . . . F = .54 not sig.
Evaluators 2-3 . . . . F = 1.33 not sig.

Summary of Reliability

$\theta$ = .71
$r_1$ = .42
$r_3$ = .68

TABLE 2B

Summary of Results for Posterior
Fillings from Evaluation Group B

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 187.00 | 174.00 | 208.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.710 | 2.522 | 3.014 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 8.5313 | 4.2656 | 21.1168 | .000 |
| Rows (Sub) | 68 | 48.9343 | .7196 | 3.5625 | .000 |
| Interaction | 136 | 27.4722 | .2020 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 2.91 not sig.
Evaluators 1-3 . . . . F = 7.70 P<.01
Evaluators 2-3 . . . . F = 20.1 P<.01

Summary of Reliability

$\theta = .85$
$r_1 = .46$
$r_3 = .72$

TABLE 2C

Summary of Results for Anterior
Preparations from Evaluation Group B

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 176.00 | 185.00 | 190.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.551 | 2.681 | 2.754 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 1.4585 | .7292 | 3.0475 | .0492 |
| Rows (Subs) | 68 | 64.3259 | .9460 | 3.9531 | .000 |
| Interaction | 136 | 32.5442 | .2393 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 1.22 not sig.
Evaluators 1-3 . . . . F = 2.96 not sig.
Evaluators 2-3 . . . . F = .38 not sig.

Summary of Reliability

$\theta = .98$
$r_1 = .49$
$r_3 = .75$

TABLE 2D

Summary of Results for Anterior Fillings from Evaluation Group B

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 193.00 | 200.00 | 215.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.797 | 2.899 | 3.116 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 3.6616 | 1.8308 | 6.8514 | .0018 |
| Rows (Subs) | 68 | 64.1807 | .9438 | 3.5321 | .000 |
| Interaction | 136 | 36.3416 | .2672 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = .66 not sig.
Evaluators 1-3 . . . . F = 6.60 P<.01
Evaluators 2-3 . . . . F = 3.60 not sig.

Summary of Reliability

$\theta = .84$
$r_1 = .46$
$r_3 = .72$

TABLE 3A

Summary of Results for Stainless Steel Crown Preparations from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 215.00 | 224.00 | 220.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 3.116 | 3.246 | 3.188 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | .5891 | .2946 | 1.6636 | .1914 |
| Rows (Subs) | 68 | 56.3550 | .8287 | 4.6806 | .000 |
| Interaction | 136 | 24.0803 | .1771 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 1.63 not sig.
Evaluators 1-3 . . . . F = .50 not sig.
Evaluators 2-3 . . . . F = .32 not sig.

Summary of Reliability

$\theta = 1.23$
$r_1 = .55$
$r_{1_3} = .78$

TABLE 3B

Summary of Results for Stainless Steel Crown Restorations from Evaluation Group A

---

# of Subjects. . . . . 69
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 221.00 | 208.00 | 210.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 3.203 | 3.014 | 3.04 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 1.4202 | .7101 | 2.4193 | .0901 |
| Rows (Subs) | 68 | 73.0984 | 1.0750 | 3.6625 | .000 |
| Interaction | 136 | 39.9167 | .2935 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 2.11 not sig.
Evaluators 1-3 . . . . F = 1.51 not sig.
Evaluators 2-3 . . . . F = .05 not sig.

Summary of Reliability

$\theta = .89$
$r_1 = .47$
$r_3 = .72$

TABLE 4A

Summary of Results for Stainless Steel Crown Preparations from Evaluation Group B

---

# of Subjects. . . . . 67
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 160.00 | 192.00 | 193.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.388 | 2.866 | 2.881 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 10.5171 | 5.2585 | 24.6561 | .000 |
| Rows (Subs) | 66 | 38.5947 | .5848 | 2.7418 | .000 |
| Interaction | 132 | 28.1523 | .2133 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 17.5 P<.01
Evaluators 1-3 . . . . F = 18.7 P<.01
Evaluators 2-3 . . . . F = .02 not sig.

Summary of Reliability

$\theta$ = .58
$r_1$ = .37
$r_3$ = .64

TABLE 4B

Summary of Results for Stainless Steel Crown Restorations from Evaluation Group B

---

# of Subjects. . . . . 67
# of Scores per Subject. . . . . 3
Sums of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 174.00 | 193.00 | 182.00 |

Means of Columns (evaluators)

| 1 | 2 | 3 |
|---|---|---|
| 2.597 | 2.881 | 2.716 |

Analysis of Variance Summary

| SOURCE | DF | SUMS OF SQUARES | MEAN SQUARES | F-RATIO | PROBABILITY |
|---|---|---|---|---|---|
| Column | 2 | 2.7158 | 1.3579 | 5.9842 | .0036 |
| Rows (Subs) | 66 | 38.8237 | .5882 | 2.5923 | .000 |
| Interaction | 132 | 29.9531 | .2269 | | |

Scheffe Post Hoc Results

Evaluators 1-2 . . . . F = 6.20 P<.01
Evaluators 1-3 . . . . F = 1.09 not sig.
Evaluators 2-3 . . . . F = 2.09 not sig.

Summary of Reliability

$\theta$ = .53
$r_1$ = .35
$r_3$ = .61

TABLE 5

Results of Student Self-Evaluation

Practical 1

# of students - 138

| | |
|---|---|
| Mean total score for students receiving bonus point | $n = 66$<br>$\bar{x} = 12.78$ |
| Mean total score for students not receiving bonus point | $n = 72$<br>$\bar{y} = 11.16$ |

$T = 6.24$ $P < .01$
$DF = 136$

Practical 2

# of students - 138

| | |
|---|---|
| Mean total score for students receiving bonus point | $n = 105$<br>$\bar{x} = 5.81$ |
| Mean total score for students not receiving bonus point | $n = 33$<br>$\bar{y} = 6.18$ |

$T = 1.12$ $P < .05$
$DF = 136$

# APPENDIX A

## AMALGAM CRITERIA CHECKLIST

## (POSTERIOR)

# AMALGAM RESTORATIONS

1. Outline Form

    1. Extension into all pits and fissures creating a bucco-lingual width slightly larger than 1mm.
    2. Depth of the preparation .5mm in dentine (1.5mm total depth)
    3. Flat pulpal floor
    4. Slight occlusal convergence of the cavity walls

2. Cavity Preparation

    1. Smooth cavo-surface margins approaching 90°
    2. Gingival seat placed below the free gingival margin in Class II restorations
    3. Buccal and lingual margins of the proximal box extended such that the lip of #6 explorer can bearly clear the margin and proximal surface of the adjacent tooth
    4. Internal angles rounded

3. Fill

    1. Anatomy resembling the original contour of the tooth with the exception of secondary anatomy
    2. Restoration ending at the cavo-surface margins
    3. Firm contact with the adjacent tooth
    4. In Class II restorations, marginal ridges, carved to the same height as the adjacent teeth

4. Polish

    1. Smoothly polished restoration
    2. No porosity or ditching of the restoration
    3. No surrounding tooth structure reduced during polish
    4. Restoration flush with the cavo surface margins

# APPENDIX B

## COMPOSITE CRITERIA CHECKLIST

## (ANTERIOR)

# COMPOSITE RESTORATIONS

A. Outline Form

1. Cavity triangular in shape with the base of the triangle towards the gingiva
2. Depth of the preparation approximately 1mm from the enamel surface
3. Buccal and lingual walls extended to bearly break contact. Labial or lingual dovetail lock is frequently indicated for retention in larger restorations. The extention of the dovetail should be at the expense of the gingival rather than incisor which might weaken that angle of the tooth

B. Cavity Preparation

1. Flat pulpal floor
2. All walls slightly divergent towards the pulpal floor to gain retention.

C. Fill

1. Anatomy resembling the original contour of the tooth
2. Restoration ending at the cavo surface margins
3. Firm contact with adjacent teeth

D. Finish

1. Smooth restoration
2. No tooth structure reduced during finish
3. Restoration flush with the cavo surface margins

# APPENDIX C

## STAINLESS STEEL CROWN CRITERIA CHECKLIST

## CRITERIA FOR EVALUATION OF S.S.C. RESTORATION

### A. Tooth Preparation

1. Occlusal reduction of 1.5mm
2. Adequate proximal reduction (with 169L bur) with 5 degree occlusal taper
3. Feather edge finish line for the prep
4. All external angles rounded

### B. Contoured and Finished Crown

1. Extension of the crown 1mm. into the gingival sulcus
2. Crown contoured proximally in order to establish firm contact
3. Crowned tooth in proper occlusal contact with the opposing teeth
4. Sufficient retention so that crown "snaps" on
5. Smoothly finished margins

## References

Abrams, R.G.: Diagnosis as an Aid to Clinical Evaluation. In R.S. Mackenzie and T.J. Harrop, Eds.: An Instructional Information Exchange for Dentistry in the United States, Volume 6, Observational Measurement of Clinical Performance. U.S. Dept. of Health, Education, and Welfare, Public Health Service, Health Resources Administration DHEWH Publication No. (HRA) 75-78, 1975, pp. 212-213.

Fuller, James L.: The Effects of Training and Criterion Models on Interjudge Reliability. Journal of Dental Education, 36: 19-22, April 1972.

Gaines, W.G., and Bruggers, H.W., and Rasmussen, R.H.: Reliability of Ratings in Preclinical Fixed Prosthodontics: Effect of Objective Scaling. Journal of Dental Education, 38: 672-675, 1974.

Hinkelman, Kenneth W., and Long, Nicholas K.: Method for Decreasing Subjective Evaluation in Preclinical Restorative Dentistry. Journal of Dental Education, 37: 13-18, September 1972.

Houpt, M.I., and Kress, G.: Accuracy of Measurement of Clinical Performance in Dentistry. Journal of Dental Education, 37: 34-46, 1973.

Irion, Arther L.: A brief History of Research on the Acquisition of Skill. In E.A. Bilodeau, Ed.: Acquisition of a Skill. New York: Academic Press, Inc., 1965

Natkin, Eugene, and Guild, Robert E.: Evaluation of Preclinical Laboratory Performance: A Systematic Study. Journal of Dental Education, 31: 152-161, June 1967.

Nedelsky, Leo: Science Teaching and Testing. Chicago: Harcourt, Brace, and World, Inc., 1965.

O'Conner, Patricia, and Lorey, Robert E.: Improving Interrater Agreement in Evaluation in Dentistry by the use of Comparison Stimuli. Journal of Dental Education, 42: 174-179, April 1978.

Ryge, G., and Snyder, M.: Evaluating the Clinical Quality of Restorations. Journal of the American Dental Association, 87 (8): 369-377, 1973.

Winer, B.J.: Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1971.

Yates, J.L., and Hembree, J.H., and McNight, J.P.: A Comparison of the Effect of Cavity Design on the Fracture Strength of Three Dental Alloy Systems. Journal of Dentistry for Children,: May-June 1976.